이동 컴퓨팅를 위한 컨텐츠 변환 기술 구현

허준회() 김형일,권문영, 이영석, 박수연
고미드
{joone(),hikim,moon,yslee,sfmelon}@gomid.com

Developing the Contents Conversion System for Mobile Computing

Joone Hur, Miso Kim, Moonyoung Kim, Brian Lee, SooYeon Park
GOMID(http://www.gomid.com)

요 약

이동형 장치로 기존 인터넷에 존재하는 정보를 접근하는데는 많은 어려움이 존재한다. 이러한 어려움을 해결하려면 기존 정보를 각 이동형 장치에 맞게 새로 작성하거나 변환해야 한다. 본 논문에서는 기존 정보를 어떻게 이동형 장치에 맞게 변환하여 제공할 수 있는 방법과 구현 내용을 소개하고 한다.

1. 서론

초고속 통신망의 보급으로 많은 사람들이 안방에서 빠르고 편리하게 인터넷을 사용하고 있다. 또한 데이터 통신이 가능한 핸드폰을 갖고 있으며 학교나 지하철과 같은 공공장소에는 무선랜이 설치되어 있어서 언제 어디서나 PDA나 핸드폰으로 인터넷에 접근할 수 있다. 하지만 PC용 웹 브라우저에 맞게 작성된 인터넷 정보를 이동형 장치에서 접근하려면 느린 전송 속도, 통신 비용, 작은 화면 등 여러 가지 제약이 많다.

본 논문에서는 기존 HTML컨텐츠 가운데 원하는 내용을 클립하여 이를 이동형 장치에서 볼 수 있는 방법을 소개하고자 한다.

2. 이동형 장치를 위한 컨텐트 제공의 어려움

근래들어 이동형 장치에서 인터넷 접근이 쉬워지면서 무선통신사업자와 CP(Contents Provider)들은 이들을 위한 전용 컨텐츠를 제공하기 시작했다.

하지만 이동형 장치를 위한 전용 컨텐츠의 양이 많지 않기 때문에 기존 컨텐츠에 접근할 필요성이 있지만 이동형장치는 휴대의 편리성과 오랜 사용시간을 위해 작은 화면과 낮은 컬러수를 지원하고 있기 때문에 기존 컨텐츠를 접근하는데 어려움이 있다.

3 컨텐츠 변환 시스템 소개

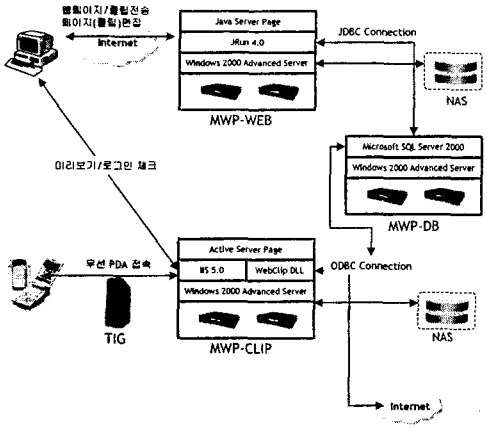
본 논문에서 구현한 컨텐츠 변환 시스템은 PC용 웹페이지 중에서 정말 필요한 부분을 미리 설정해 놓고 해당 부분만 PDA에서 볼 수 있는 시스템이다. 전체 웹페이지 중에서 우리가 실제 원하는 내용은 작은 부분일 수도 있다. 예를 들어 신문 사이트의 경우 첫 페이지에서 기사 목록만 얻어오면 신문 전체 기사를 대부분 볼 수 있다. 이때 헤드라인 기사와 각종 광고와 기사와 관계없는 링크는 선택하지 않는다.

컨텐츠 변환시스템은 전체 페이지 중 사용자가 원하는 부분에 대한 정보를 서버에 저장한다. 실제 사용자가 PDA를 이용해 원하는 웹페이지에 접근할 경우, 서버가 웹페이지에 접근하여 전체페이지에서 미리 정해놓은 부분만을 클립한 후 PDA 화면에 맞게 페이지를 재구성하여 사용자에게 서비스한다.

4 컨텐츠 변환 시스템의 구성

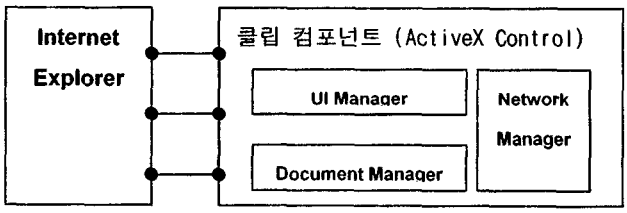
컨텐츠 변환 시스템은 크게 3부분으로 나눌 수 있다. 기존 HTML문서로부터 클립 정보를 얻는 클립콤포넌트와 클립

정보로 실제 HTML문서를 불러와 원하는 부분을 추출하는 클리핑 서버가 존재한다. 마지막으로 클립한 컨텐츠가 이동형 장치에서 잘 보일 수 있도록 HTML문서를 재구성해 주는 리포맷터가 존재한다.



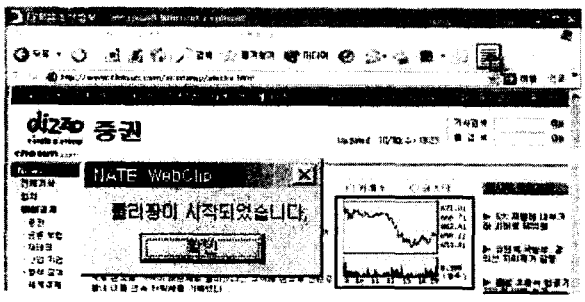
<그림1> 전체시스템구성도

4.1 클립 콤포넌트



<그림2> 클립콤포넌트 시스템 구성도

클립 콤포넌트는 마이크로소프트 인터넷 익스플로러(IE)에 나타난 웹페이지 중 원하는 부분을 TABLE, TD, TR단위로 선택할 수 있도록 하며 선택한 정보를 서버로 전송하는 역할을 담당한다.



<그림3> IE에서 클립하는 모습

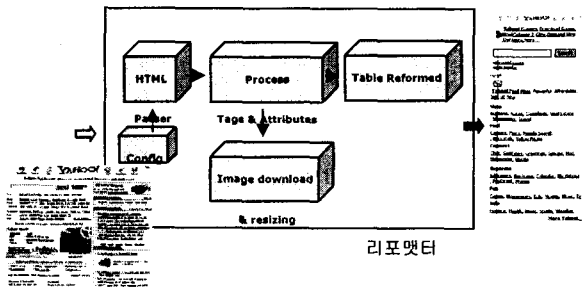
<그림2>에서 볼 수 있듯이 Document Manager는 HTML 문서 구조 분석을 처리하기 위한 모듈로서 사용자가 지정한 클리핑 영역을 HTML 문서의 구조를 분석하여 클립의 정보를 제공한다.

Network Manager는 서버와의 통신을 처리하기 위한 모듈로서 사용자 인증을 위한 사용자 아이디와 패스워드의 서버로의 송신, HTML 문서로부터 지정된 클리핑 영역에 대한 클립의 정보의 서버로의 송신, 사용자 인증에 대한 서버로부터의 응답 수신, 클립의 정보에 대한 서버로부터의 응답 수신을 담당한다.

4.2 클리핑 서버

클리핑 서버는 사용자의 요청이 있을 때, 대신 실제 웹페이지를 접근하여 전체페이지 가운데, 클립 콤포넌트에서 선택했던 영역에 해당하는 HTML문서를 클립하여 리포맷터에게 전달하는 역할을 담당하고 있다.

4.3 리포맷터



<그림4> 리포맷터 시스템 구성도

리포맷터는 크게 HTML파서, 이미지 처리기, 태그 필터링 세부분으로 나눌 수 있다. 먼저 HTML문서를 다운로드 받아 파싱을 하고 각각의 태그의 정보를 파악하여 이를 수정하고 테이블 형태를 수정한다. 이렇게 변환된 HTML은 바로 웹브라우저로 출력하거나 디스크에 저장한 후, 웹서버로 업로드할 수 있다.

5. 전체 기사 중 원하는 기사를 클리핑하는 방법

HTML문서는 트리구조로 구조화할 수 있다. 전체 문서내에 정보는 TABLE, TD, TR로 구분되어 있기 때문에 이들을 기준으로 원하는 노드를 선택할 수 있다. 노드 정보는 사용자가 직접 웹페이지로 클립을 한 후, 서버에 전달되어 저장되어 다시 사용자가 PDA로 해당 웹페이지를 접근할 때 노드 정보를 바탕으로 원하는 정보만 클리핑하여 리포맷터를 거쳐 변환된 정보를 제공할 수 있다.

6. 실험

6.1 실험환경

서버: 윈도2000서버

웹서버: MS 인터넷 인포메이션 서버

클라이언트: 싸이버뱅크 포즈(PocketPC), MS Pocket IE

통신: CDMA 모뎀

6.2 실험 결과

전체 웹페이지를 PDA에서 읽는데 걸린 시간

URL	화면 표시 시작시간(초)	읽기 완료(초)	크기 (KB)
www.daum.net	5	70	108
www.yahoo.co.kr	13	35	116
www.google.co.kr	6	19	27
www.nate.com	9	38	137
www.morning365.co.kr	60	200	924
www.joins.com	10	180	470
www.sportsseoul.com	9	157	525
www.imbc.com	15	145	340

원하는 부분만 PDA에서 읽는데 걸린 시간

URL	화면 표시 시작시간(초)	읽기 완료(초)	크기 (KB)
www.daum.net	4	12	31
www.yahoo.co.kr	3	13	22
www.google.co.kr	2	5	11
www.nate.com	3	11	23
www.morning365.co.kr	11	49	203
www.joins.com	6	23	201
www.sportsseoul.com	5	33	301
www.imbc.com	5	25	142

각각의 사이트의 초기페이지를 PDA에 탑재된 웹브라우저를 이용하여 읽는 것과 컨텐츠 변환기를 통해서 읽는 경우 걸린 시간을 비교하였다.

7. 결론

PDA에서 기존 컨텐츠를 그냥 읽는 것보다 원하는 정보를 클립핑하여 변환하여 읽는 것이 변환시간이 필요함에도 불구하고 평균 5-6배정도 빠르다는 것을 수 알 수 있었다. 이것은 전체페이지를 읽는 것이 아니라 원하는 부분만 추출하기 때문에 빠르기도 한 것이지만 서버에서 원하는 부분만 추출하는데 걸린 시간을 생각한다면 빠른 것이다. 게다가 원하는 정보만을 가로 스크롤링 없이 볼 수 있다는 것이 사용자게 편리함을 줄 수 있다.

8. 참고문헌

[1] 강성천, 이동 통신 환경에서 효율적인 웹 문서 변환에 관한 연구, 1999, 정보과학회논문지(667-669)

[2] WebClip 서비스, http://clip.nate.com